



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

How much reproducibility do we need in human and veterinary pathology?

Pospischil, A ; Folkers, G

Abstract: In diagnostic and research reports as well as text-books of human and veterinary pathology repeatability, reproducibility, inter- and intra-observer variation are mentioned rarely as a problem in preparing diagnosis from macroscopic and/or microscopic samples and discussed inconsistently. However, optimal care and restoration of health for a patient are dependent on reliability of diagnosis, therapy, prognosis and prophylaxis. This requires for all tests and procedures a maximal repeatability and reproducibility, a sensitivity and specificity of 85-95% for procedures and methodologies and a comparison of results procedures and methodologies to a gold standard. Looking at the various steps on the road to diagnosis in pathology this is influenced by a series of laboratory steps preparing tissue samples but most importantly reproducibility depends on the handling of visual information in the central nervous system of the individual diagnostician. Thus reproducibility in this context has to be divided into at least three levels: individual (epistemological, organoleptic, inter- and intra-observer variation, and formal/technological- and normative reproducibility). The aim of the present manuscript is to stimulate the reflection among the pathology experts on this most important topic.

DOI: <https://doi.org/10.1016/j.etp.2014.11.005>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-112175>

Journal Article

Accepted Version

Originally published at:

Pospischil, A; Folkers, G (2015). How much reproducibility do we need in human and veterinary pathology? *Experimental and Toxicologic Pathology*, 67(2):77-80.

DOI: <https://doi.org/10.1016/j.etp.2014.11.005>

How much reproducibility do we need in human and veterinary pathology?

Andreas Pospischil^{1,2}, Gerd Folkers²

¹Institute for Veterinary Pathology, University of Zurich, Switzerland and ²Collegium Helveticum, University of Zurich and Swiss Federal Institute of Technology, Zurich, Switzerland

Corresponding author: Andreas Pospischil, Institute for Veterinary Pathology, University of Zurich, Switzerland, Winterthurerstrasse 268, CH-8057 Zürich, email: apos@vetpath.uzh.ch

Keywords: Reproducibility-, repeatability of diagnosis, inter- and intraobserver variation,

Abstract

In pathology agreement (reproducibility) in making macroscopical and histopathological diagnosis is vital for therapy and prognosis of a patient. This, to the same extent is applicable to human and veterinary medicine. Pathology and other disciplines relying heavily on visual interpretation e.g. radiology are fields of medicine with the lowest diagnostic error rate reproducibility of diagnosis. However, in pathology this not only dependent on the handling of visual information in the central nervous system of the diagnostician and individual inter- and intra-observer variation but is influenced by a series of laboratory steps in the preparation of histopathological slides. Reproducibility thus has to be divided into at least three levels: Individual (epistemiological, organoleptic, inter- and intra-observer variation, and formal /technological- and normative reproducibility.

1. Introduction

“Veterinary pathologists achieve 80% agreement in application of WHO diagnoses to canine lymphoma.” This title of a research article by Valli (2008) implies that there is a great deal of agreement and minimal inter-observer variation among veterinary diagnostic pathologists, at

least with respect to lymphoma diagnoses. However, this is too simple an assumption in view of other investigations that report significant intra- and inter-observer variation among veterinary pathologists in grading canine cutaneous mast cell tumors (Northrup et al., 2005) or intestinal tissues (Willard et al., 2002). Neither is this problem unique to veterinary pathology; it is discussed in human pathology as well (Furness et al., 2003). Pathology and radiology relying heavily on visual interpretation are the two fields of medicine with the lowest diagnostic error rate reproducibility of diagnosis (Berner and Graber, 2008). However, this is not only dependent on the handling of visual information in the central nervous system and individual inter- and intra-observer variation but is influenced by a series of laboratory steps from a macroscopical diagnosis to taking a biopsy and finally the preparation of histopathological slides.

In general, diagnostic and therapeutic procedures, prognoses and prophylaxis in both medical and veterinary medical clinical disciplines and pathology need to be reproducible to regain patient health. But intra- and inter-observer variation is only one, probably the last link in the chain of steps (procedures and methodologies) leading from the patient (tissue sample) to diagnosis. The term reproducibility encompasses the entire set of steps and describes the certainty that by different people under varying laboratory conditions these steps can be reproduced thus describing the accuracy and reliability of a test method.

2. Requirements for the reproducibility

The requirements for the reproducibility of a test method, i.e. procedures and methodologies usually applied in making a diagnosis, are (Berner and Graber, 2008).

- High degree of overall repeatability of all procedures and methodologies applied for the entire test method;
- Sensitivity of between 85% to 95% for procedures and methodologies applied depending on the medical / veterinary speciality;
- Specificity of between 85% to 95% for procedures and methodologies applied depending on the medical / veterinary speciality;
- Comparison of procedures and methodologies applied to a gold standard.

3. Applying reproducibility to diagnostic pathology

In applying the term reproducibility to the workflow of making diagnoses in human / veterinary pathology, one realizes that there is not one type of overall reproducibility, but that the notion has to be subdivided into several levels, depending first of all on different kinds of diagnosis (macroscopic, histopathological, immunohistochemical, ultrastructural, molecular and others). These kinds of diagnosis are based on very different preparation steps and technologies applied, e.g., to a patient post mortem investigation or biopsy. To make such a diagnosis, secondly, a pathologist needs a great number of capabilities, skills, competences and experience to see, recognize and interpret image data on different magnification levels (naked eye vs. microscope). Summarizing this complex network of abilities one can differentiate reproducibility into the following levels:

- Individual (epistemological: i.e. neurobiological, gestaltpsychological (visual) and organoleptic (auditory, tactile, olfactoric) reproducibility with inter-observer and intra-observer variation (Willard et al., 2002);
- Formal /technological reproducibility (i.e. standardization of technologies and processes, measurement, analysis, ISO standards and others);
- Normative reproducibility (i.e. national and or international, e.g. WHO diagnostic guidelines or legal implications, ethics, social codes and others).

The table 1 exemplifies the necessary steps with respect to the different reproducibility levels on the way from a human or animal patient to a diagnosis creating a data set for further epidemiological evaluation.

4. Macroscopic diagnosis of a patient`s lesion

The first step in the process of making a pathological diagnosis is the naked eye evaluation of a structural difference from the norm on/in a patient organ. This leads to a macroscopic diagnosis and is primarily a visual process. In other words the eye of a pathologist is used as a measuring device, differentiating colors, brightness values, structures, sizes and movement of the lesion. The ability to differentiate between these parameters depends on the intensity of light and colors, the absolute and relative size of the lesion and training and abilities, both physical and mental, of the pathologist (https://wiki.fh-muenster.de/fb3/boesche/doku.php?id=public:mt_auge). With respect to reproducibility this procedure falls into the level of individual/epistemological reproducibility.

The interpretation of a visual impression of either artistic or scientific origin is, according to Ernst Kris and Ernst Gombrich (cited from Kandel, 2012), a creative brain process. Nature or

biology is thus comparable to the artist: Gombrich (1978) terms this process “visual discovery through art”. In particular the process of creating a macroscopic pathology diagnosis uses more than visual information. It includes organoleptic data of tactile origin, like consistency or surface temperature of a localized lesion compared to healthy surrounding areas, patient olfactory data and patient auditory events showing e.g. pain or crepitation of a fracture. Gestalt psychology summarizes this in one sentence: “The whole is more than the sum of its parts” describing the interactions of representation from visual and organoleptic information as important steps on the way to making a macroscopic diagnosis.

It was Gombrich (1978) who, following the ideas of Gestalt psychology, first realized that in the central nervous system the visual representations mentioned are processed by the basic thalamic perception system and from there via a bottom up process reach higher visual cognitive areas. In the cortical cerebral areas primary visual perceptions are further processed (compared) under consideration of previously acquired knowledge, leading to classification of information, checking of hypotheses and conclusions. Gombrich (1978) distils these processes to the sentence: “perception is based on classification of representations”, avoiding visual illusions (Kandel, 2012). A perfect example of an optical illusion is the “Necker cube” (http://en.wikipedia.org/wiki/Necker_cube).

Feldfunktion geändert

To be able to classify visual and organoleptic representations of a patient’s macroscopic lesions on the way to a diagnosis and achieve minimal intra- and inter-observer variation, a pathologist needs first of all a thorough, specialized post-graduate training. In addition, internal and external peer review by more experienced colleagues completes the gain of experience (Willard et al., 2002). Once procedures to apply internal or external peer review are standardized and formalized and written up in “standard operational procedures” (SOP) they become part of accreditation and standardization systems (<http://www.hp-management.ch/pdf/17025-grundlagen.pdf>) and are part of formal/technological reproducibility.

Feldfunktion geändert

5. Taking, processing and interpreting a biopsy

Taking a biopsy from a patient is a surgical procedure to remove a representative tissue sample from a previously diagnosed lesion for further preparation, aiming at microscopic diagnosis. All procedures involved are highly standardized. Thus with respect to reproducibility this procedure falls into the level of formal/technological reproducibility.

5.1. Fixation and embedding of tissue samples

For many years the procedures and methodologies involved are highly standardized and partially automated using specific laboratory equipment. Most of the laboratories involved in such work follow guidelines like EN / ISO 17025 (<http://www.swisstestinglabs.ch/qualitaetssicherung/iso-17025.html>) or comparable systems. Thus with respect to reproducibility, this procedure falls into the level of formal/technological and normative reproducibility.

Feldfunktion geändert

5.2. Cutting and staining of tissue sections

In most histopathological laboratories tissue sections are cut manually on microtomes, however, the methods applied are highly standardized and regulated by guidelines like EN / ISO 17025 or comparable systems. Staining is usually automated in specialized automated laboratory equipment and again regulated by guidelines like EN / ISO 17025 or comparable systems. Thus with respect to reproducibility this procedure falls into the level of formal/technological and normative reproducibility.

Microscopic reading of slides and preparation of a microscopic diagnosis and final diagnosis
To read slides, specially trained pathologists use a microscope. Although using this technical tool instead of the naked eye, visual impressions are processed in a way similar to that described earlier. In the cortical cerebral areas the primary visual perceptions (visual impression gained from the tissue section) are further processed (compared) under consideration of previously acquired knowledge, leading to classification of information (diagnoses), checking of hypothesis and conclusions (overall diagnosis).
Thus with respect to reproducibility these procedures fall into the level of individual/epistemiological reproducibility.

5.3. Quality control by internal or external peer reviews

Over decades macroscopic and histopathological diagnoses, especially of neoplastic lesions from different national or international laboratories, were hardly comparable due to local or

national schools of thought and thought styles, according to Ludwik Fleck (1936, 1979). To optimize individual / epistemological reproducibility and in particular inter-observer variation, in a first step, several different classification systems for tumors were developed to gain information on the prognosis of a specific neoplasm. This development evolved parallel to implementation of clinical tumor therapies that did not include radical surgical excision of a tumor. However, not all of the classification systems were compatible. In the 20th century, however, especially after World War II, the globalization of pathological knowledge through international communication, e.g. international case conferences, peer review and WHO activities, greatly improved the situation and subsequently led to more consensual diagnoses. Today, tumor classification is well established: the microscopic image of a tumor is classified under consideration of the structure, invasive characteristic and intensity of mitosis and markers of neoplastic cells tested immunohistochemically in grades (<http://www.who.int/classifications/icd/adaptations/oncology/en/>), and the clinical situation (National Cancer Institute, 2004) (possible metastases) of a patient is classified with respect to the histopathological diagnosis and possible prognosis into stages (National Cancer Institute, 2010) (tumor staging; Greenough, 1925). Today this procedure is complemented with information on possible alterations of a gene expression pattern Ignitiadis and Sotiriou, 2008). These developments are incorporated into existing national and or international quality control systems like EN / ISO 17025. Thus with respect to reproducibility these procedures fall into the level of formal/ technological and normative reproducibility. However, the improvements in procedures and reproducibility described have resulted in a reduction of thought styles according to Ludwik Fleck (1936, 1979).

5.4. Incorporation of diagnoses into a patient data file and coding of diagnoses for comparative reasons.

In general, reports on pathological investigations (macroscopic and or histopathologic) are sent in writing to the submitting medical/veterinary doctor who uses the information to formulate or optimize therapeutic measures. In different countries or language areas, the corresponding local language is used. At the hospital, the medical doctor and/or the pathology laboratory diagnoses the histopathologic slides, and the embedded tissue samples are stored. Reports in a patient data file are today kept predominantly in electronic form. Up to this point

we have been dealing with a diagnosis from an individual patient. However, quality control and reproducibility of diagnoses is dependent on comparison of diagnoses from a series of data from patients with similar diseases and diagnoses. To facilitate such a comparison of diagnoses from different individuals, various language areas, countries or laboratories, the WHO has developed, beginning in 2000, a coding system especially for diagnoses in oncology (ICD-O 3; <http://www.who.int/classifications/icd/adaptations/oncology/en/>). This dual classification contains codes for the topography (location) of the neoplasm on/in the patient body and codes for the histopathologic diagnosis and the malignancy of the neoplasm. Thus with respect to reproducibility these procedures fall into the level of formal/technological and normative reproducibility.

Feldfunktion geändert

5.. Epidemiological evaluation of diagnoses

The creation of a database containing a larger series of datasets from patients with neoplasia consisting of coded (ICDO-3) histopathologic diagnoses and the respective topographic location allows the evaluation of inter-observer variation on a broad basis and is able to show another aspect of formal /technological and normative reproducibility.

Histopathologic diagnoses and the location of neoplasms are checked for plausibility, e.g. is a specific neoplasm in a certain location plausible, then correlated with anonymized data of individual patients, such as age, sex, species, breed and others). Thus, large collections of datasets ("big data") can be developed and evaluated ("data mining") with respect to various questions (<http://www.unglobalpulse.org/>). Epidemiological research of this sort starts with counting and is looking for correlations between series of datasets (Mukherjee, 2010).

6. Discussion and outlook

In accredited veterinary pathology laboratories the process followed to reach macroscopic and histopathologic diagnoses in general meets the most stringent requirements for reproducibility of a test method, i.e. procedures and methodologies usually applied for making a diagnosis (Berner and Graber, 2008). In this way a pathology diagnosis is able to serve the demands of the submitting veterinary surgeon and the wellbeing of the patient awaiting optimal therapy. The availability of coded diagnoses for animal diseases, especially canine and/or feline neoplasia, including basic patient data (age, breed, sex, place of residence) in electronic relational databases with correlation to available animal population data open new

possibilities of epidemiological research beyond existing animal cancer registries (<http://www.vetcancer.dk>; Broenden et al., 2007; Dobson et al., 2002; Doll and Hill, 1950; Dorn et al., 1968a; Dorn et al., 1968b; MacVean et al., 1978; O'Brien et al., 2000; Vascellari et al., 2009).

The results of such a study will show new correlations between tumor incidence in affected species (humans, canine, feline), and age, external influences (e.g. environmental) and sex using appropriate statistical methods. Whether causalities lie behind these correlations remains to be proven - possibly experimentally. However, only the correlation of smoking and lung cancer published by Doll and Hill in 1950 [24] opened the investigation of a now proven causality of this correlation.

Acknowledgement

This manuscript was prepared at the Collegium Helveticum (<http://www.collegium.ethz.ch/index.de.html>), a laboratory of trans-disciplinary research and a joint venture of University of Zurich and ETHZ (Swiss Federal Institute of Technology Zurich), where fellows of both universities worked on the different disciplinary aspects of reproducibility from 2009 – 2014.

The authors are thankful to all co-fellows and colleagues for the inspiring atmosphere and intensive discussion of the subject.

References

- Berner ES and Graber ML. Overconfidence as a Cause of Diagnostic Error in Medicine. American Journal Medicine 2008;121, (Supplement 5A):S2-S23.
- Broenden LB, Flagstad A, Kristensen AT. Veterinary cancer registry in companion animal cancer: a review. Veterinary Comparative Oncology 2007;5, 133-144.
- Dobson JM, Samuel S, Milstein H, Rogers K, Wood JLN. Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs. Small Animal Practice 2002;43, 240-246.
- Doll R and Hill AB. Smoking and carcinoma of the lung. British Medical Journal 1950;2, 739-748.

Dorn CR, Taylor DO, Frye FL, Hibbard HH. Survey of animal neoplasms in Alameda and Contra Costa counties, California. I. Methodology and description of cases. *Journal National Cancer Institute* 1968a;40, 295-305.

Dorn CR, Taylor DO, Schneider R, Hibbard HH, Klauber MR. Survey of animal neoplasms in Alameda and Contra Costa counties, California. II. Cancer morbidity in dogs and cats from Alameda county. *Journal National Cancer Institute* 1968b;40, 307-318.

Fleck L. *The Genesis and Development of a Scientific Fact*, Chicago: University of Chicago Press. 1979.

Fleck L. *The Problem of Epistemology*, in R.S. Cohen and T. Schnelle (eds.), *Cognition and Fact - Materials on Ludwik Fleck*, Dordrecht: Reidel, 1936.

Furness PN, Taub N, Assmann KJM, Banfi G, Cosyns JP, Dorman AM, Hill CM, Kapper SK, Waldherr R, Laurinavicius A, Marcussen N, Martins AP, Nogueira M, Regele H, Seron D, Carrera M, Sund S, Taskinen EI, Paavonen T, Tihomirova T, Rosenthal R. International Variation in Histologic Grading is Large, and Persistent Feedback Does Not Improve Reproducibility. *The American Journal of Surgical Pathology* 2003;27, 805-810.

Gombrich EH. *Symbolic Images, Studies in the Art of Renaissance Vol. 2*. London, UK: Phaidon Press. 1978.

Greenough RB. Varying degrees of malignancy in cancer of the breast. *Journal Cancer Research* 1925;9, 453-462.

Ignatiadis M, Sotiriou C. Understanding the molecular basis of histologic grade. *Pathobiology* 2008;75, 104-111.

Kandel E. *The age of insight. The Quest to Understand in Art, Mind, and Brain from Vienna 1900 to the Present*, New York, NY: Random House. 2012

MacVean DW, Monlux AW, Anderson jr PS, Silberg jr SL, Roszel JF. Frequency of canine and feline tumors in a defined population. *Veterinary Pathology* 1978;15, 700-715.

Mukherjee S. *The Emperor of All Maladies: A Biography of Cancer*. New York, NY: Scribner, a Division of Simon and Schuster, Inc. 2010.

Northrup NC, Harmon BG, Gieger TL, Brown CA, Carmichael P, Garcia AP, Latimer KS, Munday JS, Rakich PM, Richey LJ, Stedman NL, Gieger TL. Variation among pathologists in histologic grading of canine cutaneous mast cell tumors. *Journal Veterinary Diagnostic Investigation* 2005;17, 245-248.

O'Brien DJ, Kaneene JB, Getis A, Lloyd JW, Swanson M. Spatial and temporal comparison of selected cancers in dogs and humans, Michigan, USA, 1964-1994. *Preventive Veterinary Medicine* 2000;47, 187-204.

Valli VEO. Veterinary pathologists achieve 80% agreement in application of WHO diagnoses to canine lymphoma. *Cancer Therapy* 2008;6, 221-226.

Vascellari M, Baioni E, Ru G, Carmionato A, Mutinelli F. Animal tumor registry of two provinces in northern Italy: incidence of spontaneous tumours in dogs and cats. *BMC Veterinary Research* 2009;5, 39-48.

Willard MD, Jergens AE, Duncan RB, Leib MS, McCracken MD, DeNovo RC, Helman RG, Slater MR, Harbison JL. Interobserver variation among histopathologic evaluations of intestinal tissues from dogs and cats. *Journal American Veterinary Medical Association* 2002; 220, 1177-1182.

Table 1: Levels of reproducibility on the way to a diagnosis in pathology

		Reproducibility levels		
Patient with single / multiple pathological lesions		Individual / epistemological, inter- / intra-observer variation	Formal / technologic	Normative
	Macroscopic diagnosis	X		
	Biopsy taken		X	X
	Fixation and embedding of tissue		X	X
	Cutting and staining of tissue sections		X	X
	Microscopic reading of slides	X		
	Microscopic diagnosis	X		
	Quality control by internal or external peer review		X	X
Final diagnosis		X		
Incorporation of diagnosis into patient data file			X	
Coding of diagnosis for comparative evaluation			X	X
Epidemiological evaluation of coded diagnoses			X	X